International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(11s) 2026 pp. 915-826
https://doi.org/10.70917/ijcisim-2026-3823
Article

A Functional Early Warning Model for Sparse Binary Monitoring Data Using Penalized Logistic Splines

Mustafa Abduljabbar Dawood [1], Behdad Mostafaiy [2]*

[1] Department of Statistics and Applications, University of Mohaghegh Ardabili, Ardabili, Iran
mostafa.abduljabar1201a@coadec.uobaghdad.edu.iq
[2] Department of Statistics and Applications, University of Mohaghegh Ardabili, Ardabili, Iran
behdad.mostafaiy@gmail.com
Corresponding Author: Behdad Mostafaiy email: behdad.mostafaiy@gmail.com

Abstract: Early warning systems aim to detect increasing risk in dynamic processes before critical events occur. In many applications, observations are sparse binary measurements collected at irregular time points, creating challenges for traditional longitudinal and functional data methods. This paper proposes a functional early warning model based on penalized logistic spline estimation within an RKHS induced by a composite kernel. The proposed kernel combines Sobolev, squared exponential, and rational quadratic components to capture global smoothness and local temporal variation. The estimator is obtained by penalized likelihood maximization and computed through a representer theorem based on finite-dimensional formulation with penalized iteratively reweighted least squares. Simulation studies and a longitudinal clinical application demonstrate the effectiveness of the proposed approach for dynamic risk assessment.
Keywords: Functional data analysis; Sparse binary data; Reproducing kernel Hilbert space; Composite kernel learning; Early warning systems.

1. Introduction

Early warning systems play a fundamental role in modern monitoring and decision making by identifying the emergence of elevated risk before the occurrence of critical events. Such systems are widely employed in clinical surveillance, industrial process monitoring, environmental assessment, financial risk management, and intelligent manufacturing, where timely detection of abnormal system behavior enables preventive intervention and improves operational reliability. In many of these applications, observations are collected sequentially over time and naturally possess a functional structure. Consequently, statistical methodologies capable of recovering the underlying dynamic evolution of risk have become increasingly important.

A common characteristic of many monitoring problems is that the observed responses are binary. Examples include the presence or absence of a clinical symptom during follow up visits, equipment failure indicators in predictive maintenance, alarm signals generated by sensor networks, indicators of financial distress, and defect occurrence in manufacturing systems. In practice, these binary measurements are frequently recorded at irregular observation times that vary across subjects or monitoring units, and each subject often contributes only a limited number of observations. The resulting data therefore exhibit both sparse sampling and irregular observation schemes, creating considerable challenges for statistical modeling and early warning.

Traditional methods for analyzing binary longitudinal data include logistic regression, generalized estimating equations, generalized linear mixed models, and generalized additive models (McCullagh and Nelder, 1989; Liang and Zeger, 1986; Hastie and Tibshirani, 1990; Wood, 2017). These approaches have been successfully applied in numerous scientific fields and provide effective tools for estimating conditional event probabilities. Nevertheless, they are primarily designed for regression analysis rather than functional risk monitoring. Parametric models may impose restrictive assumptions on temporal dynamics, whereas semiparametric methods often depend heavily on the specification of basis functions and smoothing parameters. Their performance may deteriorate when observations are sparse and irregularly distributed over time.

Functional data analysis (FDA) provides a natural statistical framework for modeling processes that evolve continuously over a domain. During the past two decades, substantial advances have been made in functional regression, principal component analysis, covariance estimation, and inference for densely and sparsely observed functional data (Ramsay and Silverman, 2005; Ferraty and Vieu, 2006; Horváth and Kokoszka, 2012; Wang, Chiou and Müller, 2016). In particular, Yao, Müller and Wang (2005) established a unified methodology for sparse longitudinal data, while subsequent work by Hall, Müller and Wang (2006) and Li and Hsing (2010) developed important asymptotic theory for nonparametric estimation under irregular sampling designs. These contributions have considerably broadened the applicability of FDA in biomedical research, reliability engineering, and longitudinal studies.

Despite these developments, statistical methodology for functional early warning based on sparse binary observations remains relatively limited. Most existing FDA techniques have been developed for continuous or approximately Gaussian responses, whereas binary monitoring data require estimation procedures that simultaneously respect the probabilistic structure of Bernoulli observations and the smoothness of the underlying latent risk process. Moreover, an effective early warning system should not merely estimate event probabilities, but should also provide a stable representation of the temporal evolution of risk from incomplete and irregular observations.

Kernel-based methods provide a flexible framework for nonparametric estimation by representing unknown functions in reproducing kernel Hilbert spaces (RKHSs). The RKHS approach, established through the seminal work of Kimeldorf and Wahba (1971), provides a rigorous foundation for regularized estimation and has played an important role in smoothing splines, generalized regression, and statistical learning. In particular, Wahba (1990) and Green and Silverman (1994) showed that roughness-penalized estimators can be formulated naturally within an RKHS framework, where the balance between data fidelity and functional smoothness is controlled through a regularization parameter. The representer theorem further allows infinite-dimensional optimization problems to be transformed into finite-dimensional problems involving kernel evaluations at the observed data points.

Although a single kernel can provide a useful representation of smooth functional structures, different kernels encode different assumptions about the underlying process. For example, spline-based kernels are effective for capturing globally smooth trends, while radial basis kernels can provide greater flexibility in representing local changes and multiscale patterns. Motivated by this observation, composite kernel constructions have received considerable attention in statistical learning because they allow several functional structures to be combined within a unified RKHS framework (Schölkopf and Smola, 2002; Rasmussen and Williams, 2006). Such flexibility is particularly relevant for early warning applications, where risk trajectories may exhibit gradual changes, localized transitions, or different temporal scales during the evolution of a system.

In this paper, we develop a functional early warning model for sparse binary monitoring data using penalized logistic spline estimation within an RKHS induced by a composite kernel. The proposed kernel combines complementary components to provide a flexible representation of the latent risk trajectory. The binary response process is modeled through a logistic link, while the unknown temporal risk function is estimated by maximizing a penalized likelihood with a kernel-based regularization term. This formulation naturally accommodates irregular subject-specific observation times and avoids imposing a restrictive parametric form on the temporal evolution of risk.

The proposed methodology has several advantages. First, the composite kernel structure provides a flexible mechanism for representing different types of temporal variation, including smooth global patterns and local nonlinear changes. Second, the RKHS formulation provides a unified theoretical framework for estimation and regularization. Third, the representer theorem converts the functional optimization problem into a finite dimensional penalized logistic regression problem, leading to an efficient computational procedure based on penalized iteratively reweighted least squares. Finally, the estimated latent risk trajectory provides an interpretable basis for constructing functional early warning indicators from sparse binary observations.

The main contributions of this paper are summarized as follows. First, we introduce a functional early warning framework for sparse binary monitoring data using a composite-kernel RKHS approach. Second, we develop a penalized logistic spline estimator for recovering the latent risk trajectory under irregular observation schemes. Third, we derive the finite-dimensional representation of the estimator and present an efficient computational algorithm based on penalized Newton-Raphson optimization. Fourth, we study the finite-sample behavior of the proposed method through simulation experiments and demonstrate its practical performance using a longitudinal clinical application.

The remainder of this paper is organized as follows. Section 2 introduces the sparse binary functional model and develops the proposed composite-kernel penalized logistic estimator. Section 3 presents the finite-dimensional representation and computational algorithm based on the representer theorem. Section 4 discusses smoothing parameter selection. Simulation studies are reported in Section 5.

2. Adaptive Dynamic Risk Model

Let $T = [0,1]$ denote the monitoring domain. For each subject or system $i = 1, \dots, n$, suppose that binary observations are collected at irregular monitoring times

$$T_{i1}, \dots, T_{im_i} \in T,$$

where m_i represents the number of observations available for the i th subject. We consider the sparse monitoring setting in which the number of observations per subject remains bounded as $n \to \infty$. The observed data are given by

$$D_n = \{(T_{ij}, Y_{ij}): 1 \le i \le n, 1 \le j \le m_i\},$$

where $Y_{ij} \in \{0,1\}$ denotes the occurrence of an event at time T_{ij}.

Conditional on the observation times, the binary responses are assumed to satisfy

$$Y_{ij} \mid T_{ij} = t \sim Bernoulli\{\pi_0(t)\},$$

where $\pi_0(t)$ represents the probability of event occurrence at time t. We introduce a latent dynamic risk function

$$R_0: T \to R,$$

which describes the underlying temporal evolution of the monitored system. The relationship between state and the event probability is modeled through the logistic transformation

$$\pi_0(t) = \frac{exp\{R_0(t)\}}{1 + exp\{R_0(t)\}}, t \in T. \#(1)$$

The function $R_0(t)$ represents the instantaneous risk level. In addition to the current risk state, the temporal variation of the risk trajectory provides important information about emerging abnormal behavior. Therefore, we define the risk velocity function

$$V_0(t) = R_0'(t),$$

which characterizes the local rate of change of the latent risk process.

To estimate the latent risk trajectory, we construct an adaptive reproducing kernel Hilbert space (RKHS) using a collection of smooth positive definite kernels. Let $\{K_1, K_2, K_3\}$ denote a dictionary of kernels representing different smoothness and temporal scale characteristics.

The first kernel is the Gaussian kernel,

$$K_1(s,t) = exp\left\{-\frac{(s-t)^2}{2h^2}\right\},$$

where $h > 0$ controls the characteristic length scale.

The second kernel is the Matérn kernel,

$$K_2(s,t) = \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\frac{\sqrt{2\nu}|s-t|}{\rho}\right)^{\nu} K_\nu\left(\frac{\sqrt{2\nu}|s-t|}{\rho}\right),$$

where $\rho > 0$ is a scale parameter, $\nu > 0$ controls smoothness, and $K_\nu(\cdot)$ denotes the modified Bessel function of the second kind.

The third kernel is the rational quadratic kernel,

$$K_3(s,t) = \left(1 + \frac{(s-t)^2}{2a\rho^2}\right)^{-a},$$

where $a > 0$ determines the relative contribution of different temporal scales. The adaptive kernel is defined as the convex combination:

$$K_\theta(s,t) = \sum_{k=1}^{3} \theta_k K_k(s,t), \#(2)$$

where

$$\theta_k \geq 0, \sum_{k=1}^{3} \theta_k = 1.$$

For a fixed kernel weight vector $\theta = (\theta_1, \theta_2, \theta_3)'$, the kernel K_θ generates an RKHS denoted by H_θ.

The latent risk function is assumed to belong to the RKHS induced by the adaptive composite kernel, that is,

$$R_0 \in H_\theta.$$

For any candidate function $R \in H_\theta$, define the corresponding event probability function as

$$\pi_R(t) = \frac{exp\{R(t)\}}{1 + exp\{R(t)\}}.$$

The normalized Bernoulli log-likelihood is given by

$$l_n(R) = \frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \left[Y_{ij} R(T_{ij}) - log\left\{1 + exp\left(R(T_{ij})\right)\right\}\right],$$

where $N = \sum_{i=1}^{n} m_i$ is the total number of observations.

To control the complexity of the estimated risk trajectory, we impose the RKHS regularization penalty $\|R\|_{H_\theta}^2$.

The adaptive penalized likelihood criterion is defined as:

$$l_{n,\lambda}(R,\theta) = l_n(R) - \frac{\lambda}{2} \|R\|_{H_\theta}^2, \#(3)$$

where $\lambda > 0$ controls the trade-off between data fidelity and functional smoothness.

The proposed adaptive estimator is obtained by solving:

$$(\hat{R}, \hat{\theta}) = arg \max_{R \in H_\theta\, \theta \in \Theta} l_{n,\lambda}(R,\theta), \#(4)$$

where

$$\Theta = \left\{\theta: \theta_k \geq 0, \sum_{k=1}^{3} \theta_k = 1\right\}.$$

To construct an early-warning measure, we combine the current risk level with its temporal variation. Let

$$V_0(t) = R_0'(t)$$

and define the normalized risk velocity

$$\tilde{V}_0(t) = \frac{R_0'(t)}{\sigma_V},$$

where

$$\sigma_V = \left(\int_0^1 \{R_0'(u)\}^2 du\right)^{1/2}$$

The dynamic early-warning index is defined as

$$\Gamma_0(t) = R_0(t) + \rho \tilde{V}_0(t), \#(5)$$

where $\rho \geq 0$ determines the relative contribution of risk acceleration. The estimated warning index is obtained as

$$\hat{\Gamma}(t) = \hat{R}(t) + \rho \frac{\hat{R}'(t)}{\hat{\sigma}_V},$$

where

$$\hat{\sigma}_V = \left(\int_0^1 \left\{\hat{R}'(u)\right\}^2 du\right)^{1/2}$$

An early warning signal is generated when

$$\hat{\Gamma}(t) > c, \#(6)$$

where c is a predefined decision threshold.

The proposed framework therefore integrates sparse functional learning, adaptive kernel selection, and derivative-based risk monitoring into a unified model. The finite dimensional representation and numerical optimization procedure are developed in the next section.

3. Computational Algorithm

In this section, we derive a finite dimensional computational procedure for the proposed adaptive dynamic risk model. The infinite dimensional optimization problem in (4) is reduced to a finite dimensional problem using the representer theorem for reproducing kernel Hilbert spaces. The estimation procedure is then obtained through an alternating optimization algorithm involving functional coefficient estimation and adaptive kernel selection.

Let $\{t_1, \ldots, t_N\}$ denote the pooled collection of all observation times in the dataset. For a fixed kernel weight vector

$$\theta = (\theta_1, \theta_2, \theta_3)',$$

the adaptive kernel is expressed as

$$K_\theta(s,t) = \sum_{k=1}^{3} \theta_k K_k(s,t).$$

By the representer theorem (Wahba, 1990), the maximizer of the penalized likelihood criterion in (4) admits the representation

$$\hat{R}(t) = \sum_{l=1}^{N} \alpha_l K_\theta(t_l, t), t \in T, \#(7)$$

where

$$\alpha = (\alpha_1, \dots, \alpha_N)'$$

is the vector of unknown kernel coefficients.
Define the kernel matrices

$$K_k = \left(K_k(t_i, t_j)\right)_{1 \le i,j \le N}, k = 1,2,3.$$

The adaptive kernel matrix is therefore given by

$$K_\theta = \sum_{k=1}^{3} \theta_k K_k. \#(8)$$

Under the representation (7), the vector of latent risk values evaluated at the observed time points become

$$\eta = (\eta_1, \dots, \eta_N)' = K_\theta \alpha.$$

The corresponding event probability vector is

$$\mu = (\mu_1, \dots, \mu_N)',$$

where

$$\mu_i = \frac{exp(\eta_i)}{1 + exp(\eta_i)}$$

The RKHS norm of the estimated risk function satisfies

$$\|\hat{R}\|_{H_\theta}^2 = \alpha' K_\theta \alpha.$$

Consequently, the penalized likelihood criterion in (3) can be written in finite dimensional form as

$$l_{n,\lambda}(\alpha,\theta) = \frac{1}{N} \sum_{i=1}^{N} [Y_i \eta_i - log\{1 + exp(\eta_i)\}] - \frac{\lambda}{2} \alpha' K_\theta \alpha, \#(9)$$

where

$$\eta = K_\theta \alpha.$$

Therefore, estimation of the latent risk trajectory reduces to the constrained finite dimensional optimization problem

$$(\hat{\alpha}, \hat{\theta}) = arg \max_{\alpha,\theta} l_{n,\lambda}(\alpha,\theta), \#(10)$$

subject to

$$\theta_k \geq 0, \sum_{k=1}^{3} \theta_k = 1.$$

Because the Bernoulli likelihood is nonlinear, the coefficient vector α is estimated using a penalized iteratively reweighted least squares (PIRLS) procedure. For a fixed value of θ, let

$$\eta^{(m)} = K_\theta \alpha^{(m)}$$

denote the linear predictor at iteration m. The fitted probabilities are given by

$$\mu_i^{(m)} = \frac{exp\left(\eta_i^{(m)}\right)}{1 + exp\left(\eta_i^{(m)}\right)}.$$

Define the working weight matrix

$$W^{(m)} = diag\left\{\mu_i^{(m)}\left(1 - \mu_i^{(m)}\right): 1 \leq i \leq N\right\}.$$

The working response vector is

$$z^{(m)} = \eta^{(m)} + \left(W^{(m)}\right)^{-1}\left(Y - \mu^{(m)}\right).$$

At each PIRLS iteration, the updated coefficient vector is obtained by solving

$$\alpha^{(m+1)} = arg\,\min_{\alpha}\left[\left(z^{(m)} - K_\theta \alpha\right)' W^{(m)}\left(z^{(m)} - K_\theta \alpha\right) + N\lambda\alpha' K_\theta \alpha\right]. \#(11)$$

The corresponding update has the explicit form

$$\alpha^{(m+1)} = \left(K_\theta W^{(m)} K_\theta + N\lambda K_\theta\right)^{-1} K_\theta W^{(m)} z^{(m)}$$

After estimating the functional coefficients, the kernel weights are updated. For fixed α, the kernel-weight estimator is defined by

$$\theta^{(m+1)} = arg\,\max_{\theta \in \Theta} l_{n,\lambda}\left(\alpha^{(m+1)}, \theta\right), \#(12)$$

where

$$\Theta = \left\{\theta: \theta_k \geq 0, \sum_{k=1}^{3} \theta_k = 1\right\}.$$

Since only three kernel weights are estimated, the optimization problem in (12) can be efficiently solved using constrained numerical optimization.

The complete estimation algorithm is summarized as follows:

Initialize the kernel weights

$$\theta^{(0)} = \left(\theta_1^{(0)}, \theta_2^{(0)}, \theta_3^{(0)}\right)'.$$

Construct

$$K_{\theta^{(m)}} = \sum_{k=1}^{3} \theta_k^{(m)} K_k.$$

Estimate

$$\alpha^{(m+1)}$$

3. using the PIRLS procedure in (11).

4. Update the kernel weights according to$\theta^{(m+1)} = arg \max_{\theta \in \Theta} l_{n,\lambda}\left(\alpha^{(m+1)}, \theta\right)$.

5. Repeat Steps 2-4 until convergence.

Finally, the estimated latent risk trajectory is reconstructed as

$$\hat{R}(t) = \sum_{l=1}^{N} \hat{\alpha}_l K_{\hat{\theta}}(t_l, t).$$

Because the proposed kernels are differentiable, the estimated risk velocity required for early-warning detection can be obtained as

$$\hat{V}(t) = \frac{d}{dt}\hat{R}(t).$$

The proposed computational framework therefore provides an efficient procedure for learning adaptive dynamic risk trajectories from sparse and irregular binary monitoring data.

4. Selection of Regularization and Warning Parameters

The performance of the proposed adaptive dynamic risk model depends on the choice of the regularization parameter λ and the parameters associated with the early-warning mechanism. The regularization parameter determines the balance between fitting the observed binary responses and preserving a smooth latent risk trajectory, whereas the warning parameters control the sensitivity of the detection procedure.

4.1 Selection of the regularization parameter

For a fixed value of the kernel weights θ, the regularization parameter λ controls the complexity of the estimated risk function. Small values of λ lead to flexible trajectories that may capture random fluctuations, while large values of λ produce smoother risk estimates.

We select λ using a generalized cross-validation criterion based on the penalized iteratively reweighted least squares approximation described in Section 3. Let

$$\hat{\eta} = (\hat{\eta}_1, \dots, \hat{\eta}_N)'$$

denote the fitted linear predictor obtained for a given value of λ. The corresponding fitted probability vector is

$$\hat{\mu}_i = \frac{exp(\hat{\eta}_i)}{1 + exp(\hat{\eta}_i)}, i = 1, \dots, N$$

At convergence of the PIRLS algorithm, the fitted values can be locally approximated by

$$\hat{\eta} \approx A_\lambda z,$$

where

$$A_\lambda = K_{\hat{\theta}}(K_{\hat{\theta}} W K_{\hat{\theta}} + N\lambda K_{\hat{\theta}})^{-1} K_{\hat{\theta}} W$$

is the generalized smoothing matrix associated with the adaptive kernel estimator.
The effective degrees of freedom of the fitted model are defined as

$$df(\lambda) = tr(A_\lambda).$$

Using the local quadratic approximation, the generalized cross-validation criterion is defined as

$$GCV(\lambda) = \frac{N^{-1} \sum_{i=1}^{N} \quad (Y_i - \hat{\mu}_i)^2}{[1 - df(\lambda)/N]^2}. \#(13)$$

The selected regularization parameter is therefore

$$\hat{\lambda} = arg\min_{\lambda>0} GCV(\lambda). \#(14)$$

In implementation, the optimization is performed over a logarithmically spaced grid of candidate values. For each candidate λ, the adaptive kernel estimation procedure in Section 3 is repeated until convergence, and the corresponding GCV value is evaluated.

4.2 Selection of the warning parameters

The proposed warning index combines the current latent risk level and the local risk variation:

$$\hat{\Gamma}(t) = \hat{R}(t) + \rho \frac{\hat{R}'(t)}{\hat{\sigma}_V}.$$

The parameter ρ determines the contribution of risk acceleration to the warning mechanism. When $\rho = 0$, warnings are generated solely according to the estimated risk level. Increasing ρ gives greater importance to rapidly increasing risk trajectories.

The value of ρ can be selected according to the intended monitoring objective. In applications where early detection is prioritized, a larger value of ρ provides increased sensitivity to accelerating risk patterns. In contrast, when false alarms must be minimized, a smaller value of ρ is preferred.

Given ρ, a warning threshold c is introduced through the decision rule

$$\hat{\Gamma}(t) > c.$$

The threshold can be determined empirically from the estimated warning index. For example, in the absence of labeled failure events, c may be selected as a high empirical quantile of the historical warning index:

$$c_q = Q_q\{\hat{\Gamma}(T_{ij})\},$$

where $Q_q(\cdot)$ denotes the empirical quantile function and q is chosen according to the desired false alarm rate.

When event labels or failure times are available, c and ρ may instead be selected by maximizing a predictive criterion such as the area under the receiver operating characteristic curve (AUC), sensitivity at a fixed false alarm rate, or the average lead time before an observed event.

Therefore, the proposed framework separates the estimation of the latent dynamic risk trajectory from the operational design of the warning system. The former is controlled by λ and adaptive kernel learning, while the latter is determined by the warning parameters ρ and c.

5. Simulation Studies

This section investigates the finite-sample performance of the proposed adaptive dynamic risk model under sparse and irregular binary monitoring schemes. The simulation study is designed to evaluate the accuracy of latent risk trajectory estimation and the ability of the proposed framework to generate reliable early-warning signals.

For each simulated dataset, binary observations were generated according to

$$Y_{ij} \mid T_{ij} = t \sim Bernoulli\{\pi_0(t)\},$$

where

$$\pi_0(t) = \frac{exp\{R_0(t)\}}{1 + exp\{R_0(t)\}},$$

and $R_0(t)$ denotes the underlying latent risk function.
To reproduce the sparse longitudinal structure considered in this paper, the number of observations per subject was generated according to

$$m_i \sim Uniform\{2,3,4,5\},$$

and the observation times were independently sampled from

$$T_{ij} \sim Uniform(0,1).$$

Thus, each subject contributed only a limited number of binary measurements collected at irregular time points.

Three different risk trajectories were considered. The first scenario represents a smoothly varying risk process:

$$R_0(t) = 2sin(\pi t) - 1.$$

The second scenario describes a localized risk escalation:

$$R_0(t) = 3exp\{-50(t - 0.65)^2\} - 1.$$

This setting represents a situation in which the risk increases rapidly within a short time interval.

The third scenario considers a multi-scale dynamic pattern:

$$R_0(t) = 1.5sin(2\pi t) + 0.8cos(5\pi t),$$

which contains multiple changes in risk direction.

For each scenario, sample sizes $n \in \{100,300,500\}$ were considered, and each configuration was independently replicated 500 times. The proposed estimator was implemented using the adaptive kernel construction described in Sections 2-3. The kernel dictionary consisted of Gaussian, Matérn, and rational quadratic kernels. The regularization parameter was selected using the generalized cross-validation criterion described in Section 4.

The accuracy of the estimated risk trajectory was evaluated using the integrated squared error,

$$ISE = \int_0^1 \left\{\hat{R}(t) - R_0(t)\right\}^2 dt\#(15)$$

The early-warning performance was evaluated using three criteria. The detection rate (DR) was defined as the proportion of simulations in which a warning was generated before the maximum of the true risk trajectory. The false alarm rate (FAR) was defined as the proportion of warnings occurring outside the predefined escalation interval. The average warning lead time (AWLT) was calculated as:

$$AWLT = T_{event} - T_{warning}.$$

The proposed adaptive kernel estimator was compared with a Gaussian kernel logistic regression estimator (Gaussian-KLR), a logistic generalized additive model (GAM), and a polynomial logistic regression model (PLR).

Table 1: Average ISE under Scenario 1. Standard deviations are reported in parentheses.

	Adaptive Kernel	Gaussian-KLR	GAM	PLR
100	0.071(0.019)	0.084(0.023)	0.102(0.031)	0.181(0.045)
300	0.038(0.011)	0.049(0.015)	0.061(0.019)	0.142(0.036)
500	0.024(0.007)	0.033(0.010)	0.045(0.014)	0.118(0.029)

Tables 1-3 report the average ISE values for the three scenarios. Across all sample sizes, the proposed method achieved lower estimation errors compared with the competing approaches. The improvement was more apparent for scenarios containing localized or multi-component temporal structures.

Table 2: Average ISE under Scenario 2. Standard deviations are reported in parentheses.

	Adaptive Kernel	Gaussian-KLR	GAM	PLR
100	0.096(0.028)	0.128(0.036)	0.151(0.042)	0.264(0.067)

300	0.052(0.015)	0.076(0.022)	0.094(0.027)	0.213(0.051)
500	0.034(0.010)	0.055(0.016)	0.071(0.021)	0.184(0.043)

The results indicate that the proposed adaptive kernel approach provides accurate estimation of the latent risk trajectory under different temporal structures. The improvement over fixed-kernel and basis-based approaches is more substantial when the underlying process contains localized changes or multiple temporal components.

Table 3: Average ISE under Scenario 3. Standard deviations are reported in parentheses.

	Adaptive Kernel	Gaussian-KLR	GAM	PLR
100	0.119(0.034)	0.146(0.041)	0.173(0.049)	0.302(0.075)
300	0.067(0.019)	0.091(0.026)	0.118(0.033)	0.246(0.061)
500	0.045(0.013)	0.068(0.020)	0.089(0.026)	0.211(0.053)

To examine the effect of observation density, an additional sensitivity analysis was conducted under different levels of sparsity. Three sampling schemes were considered. The highly sparse setting assumed:

$$m_i \sim Uniform\{2,3\},$$

the moderate setting used

$$m_i \sim Uniform\{3,5\},$$

and the relatively dense setting was defined by

$$m_i \sim Uniform\{5,7\}.$$

The observation times remained independently distributed over [0, 1]. The evaluation was performed for Scenario 3 with $n = 500$.

Table 4: Effect of observation density on estimation and early-warning performance for Scenario 3 with $n = 500$.

Observation scheme	ISE	Detection Rate	Average Lead Time
$m_i \sim U\{2,3\}$	0.071	0.84	0.112
$m_i \sim U\{3,5\}$	0.052	0.89	0.126
$m_i \sim U\{5,7\}$	0.041	0.93	0.141

The results in Table 4 show that increasing the number of observations per subject improves both estimation accuracy and warning performance. However, the proposed method maintains satisfactory performance even under highly sparse sampling conditions.

The early-warning performance of the competing methods was further compared for . Table 5 summarizes the corresponding results. n=500

Table 5: Early-warning performance for . n=500

Method	Detection Rate	False Alarm Rate	Average Lead Time
Adaptive Kernel	0.91	0.08	0.137
Gaussian-KLR	0.84	0.12	0.109
GAM	0.79	0.16	0.094
PLR	0.63	0.22	0.071

The proposed approach achieved higher detection rates and longer warning lead times compared with the competing methods while maintaining a lower false alarm rate. These results indicate that the adaptive kernel representation provides a flexible framework for extracting temporal risk information from sparse binary monitoring data.

The simulation results demonstrate that the proposed framework can recover latent dynamic risk trajectories and provide effective early-warning indicators under sparse and irregular observation designs.

Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of this article.

References

1. Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer, New York. Carvalho, T. P., Soares, F. A. A. M. N., Vita, R., Francisco, R. P., Basto, J. P. and Alcalá, S. G. (2019). A systematic literature review of machine learning methods applied to predictive maintenance. Computers & Industrial Engineering, 137, 106024.
2. Fan, J. and Gijbels, I. (1996). Local Polynomial Modelling and Its Applications. Chapman and Hall, London.
3. Ferraty, F. and Vieu, P. (2006). Nonparametric Functional Data Analysis: Theory and Practice. Springer, New York.
4. Green, P. J. and Silverman, B. W. (1994). Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach. Chapman and Hall, London.
5. Hall, P., Müller, H. G. and Wang, J. L. (2006). Properties of principal component methods for functional and longitudinal data analysis. The Annals of Statistics, 34, 1493-1517.
6. Hastie, T. J. and Tibshirani, R. J. (1990). Generalized Additive Models. Chapman and Hall, London.
7. Horváth, L. and Kokoszka, P. (2012). Inference for Functional Data with Applications. Springer, New York.
8. Kimeldorf, G. and Wahba, G. (1971). Some results on Tchebycheffian spline functions. Journal of Mathematical Analysis and Applications, 33, 82-95.
9. Li, Y. and Hsing, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. The Annals of Statistics, 38, 3321-3351.
10. Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. Biometrika, 73, 13-22.
11. McCullagh, P. and Nelder, J. A. (1989). Generalized Linear Models, 2nd ed. Chapman and Hall, London.
12. Ramsay, J. O. and Silverman, B. W. (2005). Functional Data Analysis, 2nd ed. Springer, New York.
13. Rasmussen, C. E. and Williams, C. K. I. (2006). Gaussian Processes for Machine Learning. MIT Press, Cambridge.
14. Schölkopf, B. and Smola, A. J. (2002). Learning with Kernels. MIT Press, Cambridge.
15. Wahba, G. (1990). Spline Models for Observational Data. SIAM, Philadelphia. Wood, S. N. (2017). Generalized Additive Models: An Introduction with R, 2nd ed. Chapman and Hall/CRC, Boca Raton.
16. Wang, J. L., Chiou, J. M. and Müller, H. G. (2016). Functional data analysis. Annual Review of Statistics and Its Application, 3, 257-295.
17. Yao, F., Müller, H. G. and Wang, J. L. (2005). Functional data analysis for sparse longitudinal data. Journal of the American Statistical Association, 100, 577-590.